

**Bon sens, faux sens, contresens et non-sens sens dessus dessous.
Pour une évaluation fidèle et valide de la compétence de traduction**

Groupe de recherche TRANCE (Translation Competence and Evaluation)

Philippe Anckaert

Haute Ecole Francisco Ferrer, Bruxelles et Hogeschool-Universiteit Brussel

June Eyckmans

Universiteit Gent

Daniel Justens

Haute Ecole Francisco Ferrer, Bruxelles et Institut de Recherche sur l'Enseignement des mathématiques
de Bruxelles

Winibert Segers

KU Leuven/Lessius et UCO/Université d'Angers

Résumé

En matière d'évaluation, le bon sens voudrait que l'épreuve de traduction soit le moyen le plus indiqué pour mesurer la capacité d'un individu à traduire d'une langue à une autre. La recherche empirique indique toutefois que dans bien des cas la subjectivité du correcteur et sa conception toute personnelle de ce qu'est la compétence de traduction, sont telles que son évaluation s'avère hautement infidèle et lui ôte par là-même la validité à laquelle elle semble pouvoir prétendre au nom du bon sens.

En nous basant sur des résultats d'expériences menées en Belgique depuis trois ans, nous montrerons que la méthode de correction analytique, même appliquée avec beaucoup de rigueur, n'échappe pas à d'importantes carences sur le plan de la fidélité intra- et inter-correcteur.

Ensuite, nous mettrons en évidence l'impact des fautes de sens sur le score calculé par un correcteur chevronné et consciencieux, pour 107 étudiants-traducteurs, au moyen d'une grille d'évaluation comportant dix autres critères (ajouts, calque, grammaire, orthographe, oublis, ponctuation, registre, style, vocabulaire, contresens). Nous serons amenés à reconsidérer la notion de faute de sens et donnerons, en illustrant les rapports qu'entretiennent entre eux les problèmes de sens et les autres lacunes présentes chez de jeunes traducteurs, une image en creux de la compétence de traduction elle-même.

Enfin, face aux limites de toute méthode d'évaluation (holistique, analytique ou normée), nous plaiderons pour un retour au bon sens, mais à un autre niveau que celui invoqué habituellement, à savoir par un abandon de l'approche strictement spéculative en faveur de la méthode expérimentale.

Sous son titre quelque peu badin et volontairement équivoque, cette contribution pose trois problèmes de nature différente, que nous aborderons dans l'ordre suivant.

Primo, nous traiterons du bon sens, souvent perçu comme synonyme de sens commun, qui incite les traductologues à considérer l'exercice classique de traduction comme un moyen a priori valide et donc incontestable d'évaluer la capacité d'un individu à traduire d'une langue à une autre. Nous serons ainsi amenés d'une part à préciser les notions docimologiques de fidélité et de validité dans le cadre d'une évaluation de la compétence de traduction, et d'autre part à mettre en lumière, au moyen d'exemples tirés de la recherche empirique, les pièges d'une approche de ladite évaluation basée sur le seul bon sens.

Secundo, nous nous attacherons à illustrer l'importance cardinale de la question du sens dans le processus d'évaluation, aux yeux mêmes des praticiens. Nous verrons en outre comment sont habituellement traités les faux sens et contresens, sans oublier d'examiner les rapports qu'entretiennent entre eux les problèmes de sens et les autres lacunes présentes chez des apprentis traducteurs.

Tertio, face aux limites de toute méthode d'évaluation, nous plaiderons pour la réinvestiture d'un bon sens prenant en compte la nécessité d'exercer, par voie empirique, un contrôle permanent des qualités docimologiques de nos examens de traduction, l'élaboration d'instruments de mesure fidèles et valides ouvrant en outre de nouvelles perspectives pour l'étude empirique de la nature et des stades d'acquisition de la compétence de traduction elle-même.

Bon sang/sens ne saurait mentir

Quoi de plus naturel que de recourir à l'exercice de traduction pour évaluer la compétence de traduction d'un étudiant ou d'un candidat à un poste de traducteur ? Alors que la question de la validité d'un test, c'est-à-dire sa capacité à mesurer véritablement ce que l'on veut mesurer, représente une préoccupation centrale des docimologues chargés d'évaluer la compétence linguistique, cette question n'effleure même pas l'esprit de la plupart des correcteurs lorsqu'il s'agit de se prononcer sur la compétence de traduction d'un individu. Si la notion de validité des tests de langues étrangères fait l'objet de tant de travaux depuis une cinquantaine d'années – depuis Lado (1961) jusqu'aux congrès internationaux annuels de la European Association for Language Testing and Assessment (EALTA), de l'International Language Testing Association (ILTA) ou de l'Association of Language Testers in Europe (ALTE) – qu'ils pourraient aisément remplir une bibliothèque entière, rares sont les études empiriques consacrées au sujet de la validité des tests de traduction (voir état de la question dans Gile, 2005, p. 254), tellement il semble évident que l'exercice classique de traduction ne peut mesurer rien d'autre que la compétence de traduction. C'est le bon sens même !

Or, il en va de la validité d'une épreuve de traduction comme de tout autre moyen d'évaluation : pour être valide, un instrument de mesure se doit d'être fidèle, c'est-à-dire produire des résultats identiques lorsqu'il est administré dans des conditions identiques. Que la fidélité d'un instrument ou d'un test constitue une condition nécessaire (mais non suffisante) de la validité, s'illustre facilement au moyen de la comparaison suivante. Personne ne mettra en doute la capacité d'un thermomètre médical à mesurer la température corporelle d'un individu. Mais imaginons à présent la situation suivante. En laissant un thermomètre médical dans votre bouche le temps recommandé, il indique 40 °C. Vous le laissez revenir à la température ambiante. Puis vous faites un second test où, cette fois, il indique 36 °C. Quelle décision prendre, si ce n'est de jeter le thermomètre ? Peut-on encore considérer que cet instrument mesure la température de votre corps ? Le haut degré d'imprécision voire l'absurdité des résultats fournis par l'instrument lui ôte toute validité. Il en va, hélas, de même avec une épreuve de traduction.

La stabilité des résultats d'un examen de traduction ne peut malheureusement pas s'évaluer par test-retest, entre autres à cause de l'effet d'apprentissage. Il est par contre possible de calculer la corrélation entre les notes décernées par différents correcteurs pour une même traduction. Les données récoltées et publiées par Waddington (2006) relèvent de cette démarche, même si ce chercheur s'est concentré, dans le cadre de son étude pionnière, sur le degré d'équivalence ou de divergence des méthodes de correction plutôt que sur le degré de concordance entre correcteurs. Il a notamment fait évaluer les traductions de 64 étudiants par cinq correcteurs selon deux méthodes de correction différentes. Le Tableau 1 ci-dessous reprend en partie des résultats de cette expérience (2006, p. 69) :

	METHOD A					METHOD B				
Subjects	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5
1	4,0	5,8	5,3	5,9	5,1	6,5	6,7	4,7	7,4	4,7
2	0,7	4,8	3,8	3,9	4,0	4,1	4,4	3,3	5,9	3,5
3	2,7	6,8	5,8	6,5	6,3	6,5	7,2	6,1	7,9	6,2
4	4,5	5,6	5,3	6,3	6,3	7,3	7,1	3,8	6,9	6,6
5	2,4	4,5	5,4	4,7	5,9	5,3	5,1	3,6	5,5	5,3
6	3,1	6,9	6,2	6,6	6,1	5,6	5,9	4,9	7,1	6,7
7	3,0	3,8	4,4	4,3	3,2	5,4	4,2	2,6	5,2	3,2
8	3,5	4,9	6,7	5,8	5,3	5,2	5,8	6,2	7,2	5,9
9	3,2	5,8	6,3	8,6	6,1	6,8	7,2	6,4	8,0	5,9
10	4,7	7,3	6,6	7,7	6,5	6,8	6,2	6	8,1	6,5
[...]										
MEAN	3,8	5,9	5,8	5,9	5,5	6,3	6,3	5,2	6,8	5,6

Tableau 1 :

Résultats expérience Waddington 2006

Bien que ces résultats incitent l'auteur à conclure à l'équivalence des méthodes de correction, l'examen attentif du Tableau 1 révèle que sur le plan individuel la note obtenue par un même étudiant peut varier considérablement en fonction de la méthode de correction (A ou B) et du correcteur quelle que soit la méthode. Les écarts entre les notes attribuées à une même traduction sont non seulement considérables, mais placent l'étudiant tantôt (bien) au-dessus tantôt (bien) en-deçà du seuil de réussite.

Les résultats de l'expérience menée par Rilloy, Segers, Anckaert & Eyckmans (2007) illustrent l'impact de la subjectivité du correcteur de façon encore plus spectaculaire. Ces chercheurs ont fait évaluer une même traduction par 100 correcteurs différents, tous traducteurs professionnels ou professeurs de traduction et/ou de langue, selon une même méthode. 42 % des évaluateurs ont estimé que la qualité de la traduction était suffisante pour que son auteur soit engagé dans un service de traduction, 58 % la plaçaient en-deçà du seuil de réussite. Autrement dit, la probabilité de convaincre cette centurie d'évaluateurs de la qualité de sa traduction s'est avérée proche de celle d'un pile ou face, la subjectivité de l'évaluation s'apparentant carrément à de l'arbitraire.

On voit donc qu'en matière d'évaluation de la compétence de traduction, le bon sens ou sens commun qui nous incite à considérer l'exercice classique de traduction comme étant le moyen a priori le plus indiqué, ou à tout le moins comme allant de soi, peut être mauvais conseiller. La recherche empirique indique que dans bien des cas l'imprécision des consignes de correction, la subjectivité du correcteur et sa conception toute personnelle de ce qu'est la compétence de traduction, sont telles que l'évaluation peut se révéler hautement infidèle. Ce déficit de fidélité ôte à cet instrument d'évaluation la validité à laquelle il semblait pouvoir prétendre au nom même du bon sens.

Examinons à présent le rôle joué par les fautes de sens dans la perception de la qualité d'une traduction.

Et le sens dans tout cela ?

De toute évidence, la question du sens occupe une place cardinale dans le processus d'évaluation, aux yeux mêmes des praticiens. C'est ce qui ressort entre autres des résultats d'enquête que nous avons présentés au congrès de l'EALTA, « Ethics in Language Testing and Assessment » (Anckaert, Eyckmans & Segers, 2011). Au moyen de cette enquête, menée auprès de 107 professeurs de traduction enseignant dans les dix écoles de traduction que compte la petite Belgique, nous avons essayé d'en savoir un peu plus sur les conceptions et les pratiques dominantes de nos collègues francophones et néerlandophones en matière d'évaluation. L'une des questions était la suivante :

5. Lorsque vous corrigez une traduction d'étudiant, portant sur un sujet d'ordre général, l'objet de votre évaluation est en priorité :

- ☐ la correction linguistique
- ☐ la fidélité par rapport au texte source
- ☐ la lisibilité du texte cible
- ☐ l'équivalence d'effet
- ☐ le transfert de sens
- ☐ autre :

Les résultats pour cette question, exprimés en pourcentages, figurent dans le Tableau 2 :

Correction linguistique	7 %
Fidélité au texte source	18 %
Lisibilité du texte cible	7 %
Equivalence d'effet	12 %
Transfert de sens	34 %
Autre	22 %

Tableau 2 :

Résultats question 5

Du tableau des résultats, il ressort que les pratiques ont beau diverger d'un correcteur à l'autre, il se dégage néanmoins une majorité de collègues (34 %) pour considérer que l'objet de leur évaluation est en priorité le transfert du sens. 22 % des sondés ont refusé de désigner un seul critère « prioritaire », parce

qu'ils évaluent soit en combinant les différents critères, soit en accordant plus d'importance à un ou plusieurs critères selon la nature du texte à traduire. Parmi les 22 % de collègues ayant complété la rubrique « autre », on en compte 68 %, c'est-à-dire deux tiers de ces correcteurs-là, qui intègrent toujours le transfert de sens comme élément primordial dans l'appréciation de la qualité d'une traduction.

Si la majorité des correcteurs accordent une telle importance au sens, cela devrait se manifester dans la manière dont ils évaluent des traductions, notamment dans les protocoles d'examen. Comme nous disposions de telles données, récoltées lors d'une expérience menée pour comparer la fidélité de trois méthodes d'évaluation (Eyckmans, Anckaert & Segers, 2009), nous avons décidé d'étudier l'impact des fautes de sens sur le score calculé par un correcteur chevronné et consciencieux, pour 107 étudiants-traducteurs, au moyen d'une grille d'évaluation, reprise en Annexe A, comportant dix autres critères (ajouts, calque, grammaire, orthographe, oublis, ponctuation, registre, style, vocabulaire, contresens). Cette grille d'évaluation avait été conçue et utilisée par le correcteur même dans le cadre de ses examens de traduction depuis de nombreuses années. Elle est parfaitement représentative des grilles utilisées pour l'évaluation de traductions selon la méthode dite analytique : le correcteur sanctionne les fautes selon leur nature en retranchant des points à partir d'un total donné. En l'occurrence, le total était égal à vingt.

En Annexe B, se trouve une liste illustrative des différents types de fautes relevées au moyen de cette grille d'évaluation. En ce qui concerne l'évaluation du transfert du sens, on notera les quatre choses suivantes :

1) le concepteur de la grille singularise le « contresens » des autres fautes de sens et le pénalise à un tarif double, probablement parce qu'il estime que la distorsion du sens est plus grave que dans le cas de ce qu'on appelle communément un « faux sens » ;

2) dans la catégorie des fautes de « sens », il compte le « non-sens », mais ne peut-on se demander si produire du non-sens n'est pas aussi grave que de commettre certains contresens ? Le non-sens n'est pourtant pénalisé qu'au même titre qu'un faux sens ;

3) les fautes de sens recouvrent non seulement les faux sens et les non-sens, mais également les oublis, plus les fautes de grammaire, d'orthographe et de ponctuation qui entraînent une altération du sens ;

4) la rubrique « sens » ne semble pas couvrir la notion de « sens du texte », plutôt invoquée par les correcteurs holistiques pour reprocher à un traducteur qu'il n'a pas réussi à rendre adéquatement le vouloir-dire du scripteur.

Le tableau ci-dessous illustre la répartition des catégories de fautes annotées par le correcteur. Les catégories sont présentées dans l'ordre croissant du nombre de fautes enregistrées pour chacune d'elles. La colonne de droite donne le pourcentage de fautes annotées dans chaque catégorie par rapport au total de fautes commises par l'ensemble de l'échantillon, à savoir 2 517.

	nombre de fautes	pourcentage de fautes
Ajout	0	0
Oubli	0	0
Contresens	21	0,83
Registre	34	1,35
Calque	57	2,26
Ponctuation	125	4,96
Orthographe	172	6,83
Grammaire	248	9,85
Vocabulaire	327	12,99
Style	628	24,95
Sens	905	36,07
<i>Total</i>	2 517	
<i>Moyenne</i>	23,52	

Tableau 3 :

Répartition des fautes

De l'examen du tableau, il ressort entre autres que pour le correcteur participant à l'expérience, les fautes de sens sont les plus fréquentes : 905 sur un total de fautes égal à 2 517, soit 36 %. Quel impact ont ces fautes de sens sur le score individuel des étudiants ? Le critère d'évaluation « sens » a-t-il une forte valeur prédictive par rapport au score qu'un étudiant est susceptible d'obtenir pour une traduction donnée ? Pour répondre à ces questions, nous avons procédé au calcul des corrélations entre les différentes variables, qui donne le tableau suivant :

	SCO	CON	SEN	STY	ORT	CAL	GRA	PON	VOC	REG
Score	1	-,351**	-,725**	-,307**	-,379**	-,017	-,377**	-,195*	-,534**	-,086
Contresens	-,351**	1	,370**	-,020	,187	,024	,115	,028	,293**	-,023
Sens	-,725**	,370**	1	,267**	,156	,02	,408**	,13	,390**	,116
Style	-,307**	-,020	,267**	1	,108	-,008	,283**	,165	,127	,102
Orthogr.	-,379**	,187	,156	,108	1	,083	,113	,220*	,160	-,122
Calque	-,017	,024	,018	-,008	,083	1	-,035	-,087	,048	-,080
Gramm.	-,377**	,115	,408**	,283**	,113	-,035	1	,261**	,231*	,036
Ponct.	-,195*	,028	,130	,165	,220*	-,087	,261**	1	,134	,034
Vocabul.	-,534**	,293**	,390**	,127	,160	,048	,231*	,134	1	,026
Registre	-,086	-,023	,116	,102	-,122	-,080	,036	,034	,026	1

Tableau 4 :

Corrélations variables x score

** Corrélation significative au niveau de 0,01 (bilatéral).

* Corrélation significative au niveau de 0,05 (bilatéral).

Tous les types de fautes sont corrélés négativement avec le score puisque celui-ci décroît à mesure que le nombre de fautes augmente. Ces corrélations ne sont pas toutes significatives : les fautes de registre et les calques représentent, en l'occurrence, des variables indépendantes, ayant une influence sur le score parfaitement insignifiante. Ces deux catégories de fautes ne sont d'ailleurs corrélées avec aucune autre catégorie à un niveau significatif. Elles constituent, avec le critère « ajouts » (pour lequel aucune faute n'a été enregistrée), de très pauvres indicateurs du niveau de compétence de traduction lorsqu'il s'agit d'évaluer et de discriminer des étudiants-traducteurs issus des quatre années d'études organisées en Belgique à l'époque de l'expérience. (Actuellement, les études de traducteur comportent cinq années à Bruxelles et en Wallonie.) Il n'est pas exclu que des corrélations négatives avec le score auraient pu être enregistrées pour les fautes de registre et les calques, si le groupe expérimental avait compté des traducteurs professionnels.

À l'extrême opposé, on enregistre les corrélations négatives les plus élevées avec le score pour les fautes de sens ($r=-0,725$) et celles de vocabulaire ($r=-0,534$). La valeur calculée pour le sens s'explique principalement par le nombre élevé de fautes commises dans cette catégorie-là (36,07 %), un nombre important qui s'explique lui-même par le fait que tous les oublis ont été sanctionnés par le correcteur comme des fautes de sens et que même certaines fautes de grammaire, d'orthographe et de ponctuation l'ont été également lorsqu'elles induisaient une modification du sens. Rien d'étonnant dès lors que le sens impacte tellement le score. Cette variable n'en garde pas moins une puissance explicative plutôt modeste. Selon la théorie de la régression linéaire simple, elle n'explique que 53 % de la variance (r^2) et 31 % de la mesure véritable de la variation, à savoir l'écart-type ($1-(1-r^2)^{0,5}$).

Il nous a semblé toutefois que cette faible puissance explicative de la variable « sens » était peut-être liée au mode de calcul du score choisi par le correcteur. En appliquant un retrait de points, selon la grille de correction figurant dans l'exemplier, à partir d'un total arbitraire de 20, le score maximum plafonnait à 15/20 et 43 % des étudiants obtenaient 0/20. Une des conséquences de cette « sévérité » est qu'il devient impossible, pour plus de 4 étudiants sur 10, de distinguer par exemple celui qui a commis 3 fautes de sens du camarade qui en a produit 21 ! Pour compenser cette perte d'information, nous avons recalculé les scores selon les mêmes critères d'évaluation que ceux utilisés par le correcteur expérimental, mais en opérant cette fois-ci les retraits de points à partir d'un total correspondant au nombre le plus élevé de fautes commises par un même étudiant de l'échantillon. Ce nombre maximum était 53, correspondant à un nombre maximum de points égal à 40 (le chiffre rond relève du pur hasard). Les notes obtenues ont été

ensuite ventilées sur 20 pour permettre la comparaison avec les scores obtenus selon la première méthode de calcul. Dans ces conditions, la meilleure note s'élève à 17,5/20 et un seul étudiant se voit attribuer la note 0.

Nous avons ensuite recalculé les corrélations à partir des notes pondérées. Les résultats de ces calculs figurent dans le tableau suivant :

	NOT	CON	SEN	STY	ORT	CAL	GRA	PON	VOC	REG
Note	1	- ,464**	- ,906**	- ,426**	- ,309**	-,132	- ,545**	-,273*	- ,620**	-,120
Contresens	- ,464**	1	,370**	-,020	,187	,024	,115	,028	,293**	-,023
Sens	- ,906**	,370**	1	,267**	,156	,018	,408**	,130	,390**	,116
Style	- ,426**	-,020	,267**	1	,108	-,008	,283**	,165	,127	,102
Orthogr.	- ,309**	,187	,156	,108	1	,083	,113	,220*	,16	-,122
Calque	-,132	,024	,018	-,008	,083	1	-,035	-,087	,048	-,080
Gramm.	- ,545**	,115	,408**	,283**	,113	-,035	1	,261**	,231*	,036
Ponct.	- ,273**	,028	,130	,165	,220*	-,087	,261**	1	,134	,034
Vocabul.	- ,620**	,293**	,390**	,127	,160	,048	,231*	,134	1	,026
Registre	-,120	-,023	,116	,102	-,122	-,080	,036	,034	,026	1

Tableau 5 :

Corrélations variables x note

** Corrélation significative au niveau de 0,01 (bilatéral).

* Corrélation significative au niveau de 0,05 (bilatéral).

L'examen de ces résultats révèle que les calques et les fautes de registre demeurent des variables indépendantes, car les corrélations avec les notes ne sont pas significatives. Ces deux catégories de fautes se caractérisent donc par une valeur prédictive quasi nulle pour la détermination du niveau de la compétence de traduction de l'échantillon de population étudié. La raison en est simple : les meilleurs traducteurs (ayant obtenu un score supérieur ou égal à 10/20 selon le premier mode de calcul, soit seulement 14 % de l'échantillon) pratiquent le calque ou commettent des fautes de registre avec autant de régularité que les traducteurs moins bons. Autrement dit le pouvoir discriminant des passages concernés par ces deux types de fautes est nul dans le cas du calque et extrêmement faible pour le registre.

Le deuxième mode de calcul de la note donne, par contre, des indices de corrélation beaucoup plus élevés pour les fautes de sens ($r=-0,906$), de vocabulaire ($r=-0,620$) et de grammaire ($r=-0,545$). Lorsqu'on opère une mesure plus fine des différences de niveau entre étudiants, la combinaison de ces trois facteurs explique 93,3 % de la variance des notes, soit plus de 74 % de l'écart-type. Le sens seul a une puissance explicative de la variance égale à 81 %. La capacité à transférer le sens d'une langue à l'autre semble donc bien être l'indicateur le plus puissant de la compétence de traduction lorsqu'il s'agit d'évaluer des apprentis traducteurs.

Enfin, les fautes de style ont beau représenter un quart des fautes relevées par le correcteur expérimental (24,95 %) et donc avoir une influence non négligeable sur la note ($r=-0,426$), elles ne corrélaient à un niveau significatif qu'avec le sens et la grammaire, et ce de façon très modeste (respectivement 0,267 et 0,283). Il ne semble pas y avoir de lien avec les autres catégories de fautes, comme si le style était déconnecté des autres aspects de la compétence de traduction. Cette variable apparaît-elle comme indépendante parce que le critère du style a été utilisé comme une catégorie fourre-tout qui permet de labelliser les fautes n'entrant pas dans les autres catégories ou parce que l'appréciation du style relève, encore plus que l'évaluation des autres aspects de la compétence de traduction, de la subjectivité du correcteur ? Les quelques exemples suivants sont éclairants à ce sujet.

1) [...] om steeds meer volk naar haar producten te lokken.

([...] in order to draw a larger public to the products.)

[...] pour attirer de plus en plus de *population vers ses produits.

2) Geroutineerde verkopers weten dat: [...].

(Practised salesmen know that: [...].)

Les *vendeurs habitués le savent : [...].

3) Met andere woorden: daarbuiten zijn er geen commerciële bedoelingen, wil niemand u verleiden, alleen informeren en amuseren.

(In other words: apart from that there are no commercial intentions, nor does anyone aim to seduce you, only inform and amuse you.)

*En d'autres mots : *en dehors de cela, il n'y a aucun but commercial, personne ne veut vous séduire, seulement vous informer et vous divertir.

4) Zij geven immers niet de indruk die sterke emoties te gebruiken om iets te verkopen, maar om iets te vertellen, iets aan te bieden waar het publiek naar verlangt: ontspanning, spanning, informatie, passie, avontuur, een vlucht uit de realiteit.

(They do not give the impression to use these emotions to sell something but to tell you something, to offer something that the public desires: leisure, tension, information, passion, adventure, an escape from reality.)

Ils ne donnent du moins pas l'impression d'utiliser les émotions fortes pour vendre, mais pour raconter, pour proposer quelque chose que le public désire : détente, suspense, information, passion, aventure, *fuir la réalité.

Elles ne donnent pas l'impression de faire appel à ces émotions fortes pour vendre, mais bien pour raconter, pour répondre aux attentes du public : divertissement, suspense, information, passion, aventure, *un instant de rêve.

Il n'est pas exclu que d'autres correcteurs considéreraient spontanément que les traductions 1 et 2 constituent des fautes de vocabulaire plutôt que de style. Les fautes de traduction (3) présentent toutes les caractéristiques du calque. Quant à 4, il semble difficile de trouver un consensus sur le caractère erroné des traductions rejetées, ce qui pose les questions de la norme, du purisme et de la littéralité : quelle traduction est acceptable ou non ?

Le problème de la grille d'évaluation utilisée est que les catégories de fautes ne sont pas mutuellement exclusives : non seulement les oublis sont traités comme des fautes de sens (et les ajouts auraient dû l'être par symétrie), mais les fautes de grammaire, d'orthographe et de ponctuation le sont également si ces fautes altèrent le sens. Étant donné que ce degré d'altération du sens fait l'objet d'une appréciation subjective et que les facteurs de pondération varient du simple au double, une même traduction peut donner lieu à des scores sensiblement différents si elle est évaluée par des correcteurs différents utilisant pourtant la même grille. L'analyse détaillée de la catégorisation des fautes par le correcteur expérimental (Anckaert, Eyckmans & Segers, 2010) révèle qu'un même correcteur, aussi consciencieux soit-il, peut changer d'avis pour une même faute, lorsqu'il est amené à évaluer une centaine de copies. La frontière entre d'une part le style et d'autre part le vocabulaire, la grammaire ou les calques, entre le calque et le vocabulaire, entre la grammaire et l'orthographe ne semble pas toujours claire pour le concepteur de la grille d'évaluation lui-même.

La difficulté à catégoriser les fautes de traduction, que ce soit pour le concepteur d'une grille d'évaluation ou pour d'autres correcteurs, la variabilité du degré de sévérité d'un correcteur à l'autre voire pour un même correcteur, sans compter les petites défaillances, bien humaines et toute personnelles, de l'évaluateur confronté à de longues corrections, tous ces facteurs peuvent saper complètement la stabilité de l'évaluation, la rendant peu fidèle, ce qui à son tour jette le doute sur sa validité.

Quel bon sens ?

Parmi les méthodes d'évaluation de la compétence de traduction, on peut en distinguer trois selon la place plus ou moins grande qu'elles accordent à la subjectivité – et au bon sens ? – du correcteur. Il y a tout d'abord l'évaluation holistique consistant en un jugement synthétique global, où l'examineur apprécie les différents aspects et la qualité de la traduction de manière intuitive voire impressionniste. Pour pallier les limites de cette approche globale, d'autres ont imaginé une évaluation analytique basée sur un ensemble de critères censés refléter les aspects pertinents du construit à évaluer. Les résultats d'expériences menées depuis quelques années montrent hélas que la méthode de correction analytique, même appliquée avec beaucoup de rigueur, n'échappe pas à d'importantes carences sur le plan de la fidélité intra- et inter-correcteurs. C'est pourquoi, quelques chercheurs (Eyckmans, Anckaert & Segers,

2009) proposent désormais une évaluation normée de la compétence de traduction, qui ne se base ni sur la nature ni sur le degré de gravité des fautes de traduction, mais sur leur capacité à distinguer les traducteurs compétents, caractérisés par un faible nombre de fautes, des moins compétents, caractérisés eux par un nombre élevé de fautes, quelles qu'elles soient. L'évaluation normée classe donc les traducteurs les uns par rapport aux autres en fonction de leur capacité à traduire correctement des segments de texte dont on a établi par analyse préalable un pouvoir discriminant satisfaisant. Cette méthode permet de neutraliser la subjectivité du correcteur individuel, de calculer un indice de fidélité (par exemple, l'alpha de Cronbach) comme pour les tests de langue, mais elle suppose la mise en œuvre de gros moyens qui ne sont guère envisageables que pour des examens à forts enjeux ou pour des travaux de recherche scientifique. De plus, il s'est avéré que l'évaluation normée, pas plus que les méthodes holistique et analytique, n'échappe à l'effet texte, même si le test utilisé obtient à l'analyse un niveau de fidélité très élevé ($\alpha > ,90$) (Eyckmans, Anckaert & Segers, 2010).

Face aux limites de toute méthode d'évaluation, qu'elle soit holistique, analytique ou normée, nous plaidons pour la réinvestiture d'un bon sens prenant en compte la nécessité d'exercer, par voie empirique, un contrôle permanent des qualités docimologiques de nos examens de traduction. Concrètement, nous proposons de recourir aux méthodes et aux instruments de la statistique, de la psychométrie et de la docimologie appliquée aux langues pour contrôler la fidélité des épreuves, d'une part, et d'exploiter les acquis de la traductologie pour s'en assurer la validité de contenu d'autre part.

Par ailleurs, l'élaboration d'instruments de mesure fidèles et valides ouvre de nouvelles perspectives pour l'étude empirique de la nature et des stades d'acquisition de la compétence de traduction elle-même. C'est ce que nous avons illustré en mettant en lumière les rapports qu'entretiennent entre eux les problèmes de sens et les autres lacunes présentes chez de jeunes traducteurs. Faute de modèle théorique satisfaisant, la généralisation de ce type d'études empiriques, menées avec d'autres langues, avec d'autres types de textes et avec d'autres échantillons de population, devrait nous permettre d'obtenir une image en creux de la compétence de traduction elle-même. Bref, le bon sens dans notre discipline ne consisterait-il pas en priorité à s'approprier les méthodes de recherche et les instruments de mesure des sciences dites exactes ? La réponse tombe sous le sens.

Références

- Anckaert, Ph., Eyckmans, J. & Segers, W. (2010). Holistische, analytische en normgerelateerde evaluatiemethoden. Een vergelijkend onderzoek. Hoe betrouwbaar is een nauwkeurige evaluator? *Séminaire de recherche à la Hogeschool-Universiteit Brussel*. Bruxelles : HUB.
<https://lirias.lessius.eu/handle/123456789/6261>
- Anckaert, Ph., Eyckmans, J. & Segers, W. (2011). Ethical Problems and Solutions in Testing Translation Competence. *EALTA « Ethics in Language Testing and Assessment »*. Sienne : Università per Stranieri.
- Eyckmans, J., Anckaert, Ph. & Segers, W. (2009). The perks of norm-referenced translation evaluation. Dans Cl.V. Angelelli & H.E. Jacobson (dir.), *Testing and Assessment in Translation and Interpreting Studies* (pp. 73-93). Amsterdam/Philadelphie : John Benjamins.
- Eyckmans, J., Anckaert, Ph. & Segers, W. (2010). Translation assessment within a common European framework? *EALTA « Collaboration in Language Testing and Assessment »*. La Haye : Cito.
<http://lirias.lessius.eu/handle/123456789/4969>
- Gile, D. (2005). La traduction. La comprendre, l'apprendre. Paris : PUF.
- Lado, R. (1961). *Language Testing. The Construction and Use of Foreign Language Tests*. Londres : Longmans.
- Rillof, P., Segers, W., Anckaert, Ph. & Eyckmans, J. (2007). Is de Tienvoorbeeldenmethode een betrouwbare evaluatiemethode? *Postdoctoraal onderzoek vertaalwetenschap*. Gand : HoGent.
<http://lirias.lessius.eu/bitstream/123456789/3132/2/FVTO.pdf>
- Waddington, Chr. (2006). Measuring the effect of errors on translation quality. *Lebende Sprachen, LI(2)*, 67-71.

Annexe

Annexe A : Consignes aux correcteurs utilisant la méthode analytique

Avant de corriger, veuillez prendre connaissance des consignes qui ont été fournies aux étudiants et qui sont reprises au-dessus du texte à traduire.

La méthode analytique consiste à corriger selon la grille d'évaluation reprise ci-dessous. Cette manière de procéder implique que le correcteur souligne chaque faute (de langue ou/et de traduction) et indique dans la marge la nature de la faute sous forme d'abréviation (par exemple, « S » pour sens, « CS » pour contresens, « GR » pour grammaire, etc.). Ensuite, il retrace du total de 20 un nombre de points par faute (par exemple, -1 pour une faute S, -2 pour un CS, -0,5 pour une faute GR, etc.).

SENS	Toute altération du sens dénotatif : informations erronées, non-sens... J'inclus dans cette rubrique les oublis importants, c'est-à-dire faisant l'impasse sur une information d'ordre sémantique.	-1
CONTRESENS	L'étudiant affirme le contraire de ce que dit le texte : information présentée de manière positive alors qu'elle est négative dans le texte, confusion entre l'auteur d'une action et celui qui la subit...	-2
VOCABULAIRE	Choix lexical inadapté, collocation inusitée...	-1
CALQUE	Utilisation d'une structure littéralement copiée et inusitée en français.	-1
REGISTRE	Selon la nature du texte ou la nature d'un extrait (par exemple, un dialogue) : traduction trop (in)formelle, trop recherchée, trop simpliste...	-0,5
STYLE	Lourdeurs, répétitions maladroitement, assonances malheureuses...	-0,5

GRAMMAIRE	Erreurs grammaticales en français (par exemple, mauvais accord du participe passé, confusion masculin/féminin, accords fautifs...) + mauvaise compréhension de la grammaire du texte original (par exemple, un passé rendu par un présent...) et pour autant que ces erreurs ne modifient pas en profondeur le sens.	-0,5
OUBLIS	Voir SENS	-1
AJOUTS	Ajout d'informations non contenues dans le texte (sont exclus de ce point les étoffements stylistiques).	-1
ORTHOGRAPHE	Erreurs orthographiques, pour autant qu'elles ne modifient pas le sens.	-0,5
PONCTUATION	Oubli ou utilisation fautive de la ponctuation. Attention : l'oubli, par exemple, d'une virgule induisant une compréhension différente du texte, est considéré comme une erreur de sens.	-0,5

Annexe B : Exemples des différents types de fautes de traduction

Sens

Zonder zich erom te bekommeren of de koper achteraf tevreden zal zijn over het verkochte product.
(Without worrying about the fact if the consumer will be satisfied afterwards with the acquired product.)

Ils ne se soucient même pas de savoir si l'acheteur sera satisfait * du produit acheté.

Contresens

Voor wie iets te verkopen heeft, is dat een groot psychologisch nadeel.

(For someone who has something to sell, this is a great psychological disadvantage.)

C'est un grand *avantage psychologique pour celui qui a quelque chose à vendre.

Vocabulaire

Ze pakt hierbij uit met de sterkste emoties om steeds meer volk naar haar producten te lokken.

(The strongest emotions are employed in order to get more people interested in the products.)

Elle *dégage les plus fortes émotions pour attirer davantage de consommateurs.

Calque

Dat publiek wordt vervolgens verkocht aan het bedrijfsleven, [...].

(This public is subsequently sold to the business world, [...].)

Ce public est ensuite vendu à la *vie d'entreprise, [...].

Registre

De media verspreiden een negatief beeld van reclame.

(The media diffuse a negative image of advertising.)

Les médias diffusent une vision négative de la *pub.

Style

Voor wie iets te verkopen heeft, is dat een groot psychologisch nadeel.

(For someone who has something to sell, this is a great psychological disadvantage.)

*Ainsi, cela constitue un grand inconvénient psychologique pour celui qui a l'intention de vendre quelque chose.

Grammaire

Terwijl de producten uit de amusementsindustrie gratis gepromoot worden door de journalisten in een redactionele context, [...].

(While the products from the entertainment industry get free publicity through journalists work in an editorial context, [...].)

Alors que les produits issus de l'industrie du divertissement sont vantés gratuitement par *un journaliste dans un contexte rédactionnel, [...].

Oublis

Pas d'occurrence dans le corpus, selon le correcteur.

Ajouts

Pas d'occurrence dans le corpus, selon le correcteur.

Orthographe

Ze komen dan immers minder betrouwbaar over, en geven de indruk maar in één zaak geïnteresseerd te zijn: geld.

(For they give the impression of being less reliable and only interested in one thing: money.)

Ils apparaissent donc moins fiables et donnent l'impression de n'être *intéressé que par une seule chose : l'argent.

Punctuation

Het gaat hier dus om commerciële communicatie, met even commerciële bedoelingen als de reclame.

(We are dealing with commercial communication that shares the same commercial goals as advertising.)

On a donc ici affaire à de la communication commerciale * dont les objectifs sont aussi commerciaux que ceux de la publicité.

